Lecture Notes: Hyperplanes as Machine Learning Classifiers

Learning Objectives

By the end of this lecture, you will be able to:

1. Express the equation of a line in any dimension using parametric form
2. Derive and interpret the equation of a hyperplane using normal vectors
3. Connect the hyperplane equation to the fundamental linear ML model
4. Understand how the weight vector and bias geometrically define a classifier

---

1. The Geometric View of Classification

Motivation

At its core, a classifier's job is to **separate points** belonging to different classes. Consider the cat-brain problem from Chapter 1: given inputs (hardness, sharpness), the model must decide "threat" vs. "not threat."

Geometrically, each input is a **point** in feature space:

- 2 features → point in 2D plane
- 3 features → point in 3D space
- n features → point in n-dimensional space

Key Insight: A linear classifier works by finding a boundary (line, plane, or hyperplane) that separates classes.

---

2. Multidimensional Line Equation

Why We Need a New Formulation

The familiar high school equation y = mx + c doesn't generalize well to higher dimensions. We need a formulation that works in **any** dimension.

Parametric Line Equation

Definition: A line joining two points $\vec{a}$ and $\vec{b}$ can be expressed as:

$$\vec{x} = \vec{a} + \alpha(\vec{b} - \vec{a}) = (1 - \alpha)\vec{a} + \alpha\vec{b}$$

where $\alpha \in \mathbb{R}$ is a parameter.

Geometric Interpretation

To reach any point on the line:

1. **Start** at point $\vec{a}$
2. **Travel** along direction $(\vec{b} - \vec{a})$
3. **Distance** determined by parameter α

Value of α	Location on Line
$\alpha = 0$	At point $\vec{a}$
$\alpha = 1$	At point $\vec{b}$
$0 < \alpha < 1$	Between $\vec{a}$ and $\vec{b}$
$\alpha < 0$	Beyond $\vec{a}$ (opposite side from $\vec{b}$)
$\alpha > 1$	Beyond $\vec{b}$ (opposite side from $\vec{a}$)

Note on Linear vs. Convex Combinations

The coefficients $(1 - \alpha)$ and α always sum to 1:

$$(1 - \alpha) + \alpha = 1$$

However, this is a **convex combination** only when $0 \leq \alpha \leq 1$ (both weights non-negative). When α is outside [0,1], one coefficient becomes negative:

- α = 2: weights are -1 and 2 (sum to 1, but not convex)
- α = -0.5: weights are 1.5 and -0.5 (sum to 1, but not convex)

---

3. Multidimensional Plane (Hyperplane) Equation

Step 1: What Defines a Plane?

A plane has a special property: there exists a direction called the **normal** $\hat{n}$ that is perpendicular to the plane surface at every point.

Think of a table top. The direction pointing straight up from the table is the normal. No matter where you stand on the table, "up" is the same direction.

Step 2: Vectors That Lie Within the Plane

Consider two points **on the plane**: a fixed reference point $\vec{x}_0$ and any other point $\vec{x}$.

The vector connecting them is:

$$\vec{x} - \vec{x}_0$$

Critical observation: This difference vector lies **entirely within the plane** because both its endpoints are on the plane.

```
         n̂ (normal - points OUT of plane)
         ↑
         |
         |
    ─────●────────●──────────  ← The plane
        x₀    →    x
           (x - x₀) lies IN the plane
```

Step 3: The Perpendicularity Condition

Since $\hat{n}$ is perpendicular to the plane, it must be perpendicular to **any vector lying in the plane**.

Therefore:

$$\hat{n} \perp (\vec{x} - \vec{x}_0)$$

Perpendicular vectors have zero dot product:

$$\hat{n} \cdot (\vec{x} - \vec{x}_0) = 0$$

This is the defining equation of the plane!

Step 4: Dot Product Notation

The dot product can be written in two equivalent ways:

Notation 1 — Dot operator:

$$\hat{n} \cdot \vec{x} = n_1 x_1 + n_2 x_2 + \cdots + n_k x_k$$

Notation 2 — Transpose and matrix multiplication:

$$\hat{n}^T \vec{x} = \begin{bmatrix} n_1 & n_2 & \cdots & n_k \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = n_1 x_1 + n_2 x_2 + \cdots + n_k x_k$$

These are identical operations!

$$\boxed{\hat{n} \cdot \vec{x} = \hat{n}^T \vec{x}}$$

The transpose notation is preferred in ML because:

- Matrix multiplication rules apply (easier to chain operations)
- Consistent with code: `torch.matmul(n.T, x)` or `n.T @ x`
- Generalizes when $\vec{x}$ becomes a matrix of multiple data points

Important: The textbook's equation 2.9 states $\vec{w}^T \vec{x} = \vec{x}^T \vec{w}$. This commutativity **only holds for vectors** (both produce the same scalar). For general matrices, $A^T B \neq B^T A$.

Step 5: Expanding the Plane Equation

Starting from:

$$\hat{n} \cdot (\vec{x} - \vec{x}_0) = 0$$

Using transpose notation:

$$\hat{n}^T(\vec{x} - \vec{x}_0) = 0$$

Distributing:

$$\hat{n}^T\vec{x} - \hat{n}^T\vec{x}_0 = 0$$

Step 6: Identifying the Bias Term

The term $\hat{n}^T\vec{x}_0$ is a **scalar constant** because:

- $\hat{n}$ is fixed (the plane has one specific normal direction)
- $\vec{x}_0$ is fixed (a specific known point on the plane)
- Their dot product yields a single number

Example:

$$\hat{n} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \vec{x}_0 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

$$\hat{n}^T\vec{x}_0 = 2(0) + 3(2) = 6 \quad \text{(just a number)}$$

Define the bias: $b = -\hat{n}^T\vec{x}_0$

Then the plane equation becomes:

$$\boxed{\hat{n}^T\vec{x} + b = 0}$$

---

4. The Critical Connection to Machine Learning

The Linear Model Revisited

Recall from Chapter 1, the simplest ML model computes:

$$y = w_0 x_0 + w_1 x_1 + \cdots + w_n x_n + b = \vec{w}^T \vec{x} + b$$

For classification, we use the **decision boundary** where output equals zero:

$$\boxed{\vec{w}^T \vec{x} + b = 0}$$

The Geometric Revelation

Comparing the two equations:

Hyperplane Equation	ML Decision Boundary
$\hat{n}^T \vec{x} + b = 0$	$\vec{w}^T \vec{x} + b = 0$

Correspondence:

- **Weight vector** $\vec{w}$ ↔ **Normal direction** $\hat{n}$
- **Bias** b ↔ **Position parameter** $(-\hat{n}^T \vec{x}_0)$

What Training Actually Learns

During training, we are learning:

1. **Weights** $\vec{w}$ → The **orientation** (tilt) of the separating hyperplane
2. **Bias** b → The **position** (location) of the hyperplane in space

This is the fundamental geometric meaning of a linear classifier!

---

5. Proof: The Weight Vector Is Normal to the Decision Boundary

Direct Proof

Suppose we have two points $\vec{x}_1$ and $\vec{x}_2$ that are **both on the decision boundary**:

$$\vec{w}^T\vec{x}_1 + b = 0$$

$$\vec{w}^T\vec{x}_2 + b = 0$$

Subtracting the second from the first:

$$\vec{w}^T\vec{x}_1 - \vec{w}^T\vec{x}_2 = 0$$

Factor out:

$$\vec{w}^T(\vec{x}_1 - \vec{x}_2) = 0$$

Interpretation: The dot product of $\vec{w}$ with $(\vec{x}_1 - \vec{x}_2)$ is zero.

Since $(\vec{x}_1 - \vec{x}_2)$ is a vector connecting two points on the boundary, it lies **within** the boundary surface.

Since this holds for **any** two points on the boundary, $\vec{w}$ is perpendicular to **every** direction within the boundary.

Conclusion: $\vec{w}$ is the normal vector to the decision boundary. ∎

Concrete 2D Example

Consider the line $2x_1 + 3x_2 - 6 = 0$

Here $\vec{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $b = -6$.

Find two points on the line:

- Point A: Let $x_1 = 0 \Rightarrow 3x_2 = 6 \Rightarrow x_2 = 2$. So $\vec{x}_A = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$
- Point B: Let $x_2 = 0 \Rightarrow 2x_1 = 6 \Rightarrow x_1 = 3$. So $\vec{x}_B = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$

Vector along the line:

$$\vec{x}_B - \vec{x}_A = \begin{bmatrix} 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Check perpendicularity:

$$\vec{w} \cdot (\vec{x}_B - \vec{x}_A) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ -2 \end{bmatrix} = 2(3) + 3(-2) = 6 - 6 = 0 \checkmark$$

The dot product is zero, confirming $\vec{w}$ is perpendicular to the line!

---

6. Classification Using the Hyperplane

The Sign Test

For any input point $\vec{x}$, compute:

$$f(\vec{x}) = \vec{w}^T \vec{x} + b$$

Classification rule:

- If $f(\vec{x}) > 0$: Point is on the **positive side** → Class A
- If $f(\vec{x}) < 0$: Point is on the **negative side** → Class B
- If $f(\vec{x}) = 0$: Point is exactly **on** the hyperplane (boundary)

Geometric Interpretation

```
                    ↑ w⃗ (normal direction)
                    |
                    |
     POSITIVE SIDE  |
     (w⃗ᵀx⃗ + b > 0)  |
                    |
 ═══════════════════●═══════════════════  Decision boundary
                   x₀                       (w⃗ᵀx⃗ + b = 0)
     NEGATIVE SIDE
     (w⃗ᵀx⃗ + b < 0)       →  direction along boundary
                            (perpendicular to w⃗)
```

The weight vector $\vec{w}$:

1. Points **perpendicular** to the decision boundary
2. Points **toward** the positive classification region
3. Its **magnitude** affects the steepness of the transition between classes

---

7. Distance from a Point to the Hyperplane

Signed Distance Formula

For a hyperplane $\vec{w}^T\vec{x} + b = 0$ and a point $\vec{p}$:

$$\text{signed distance} = \frac{\vec{w}^T\vec{p} + b}{\|\vec{w}\|}$$

Why This Matters

1. **Confidence measure**: Points farther from the boundary are classified more confidently
2. **Support Vector Machines**: Maximize the margin (distance) from the boundary
3. **Model calibration**: Distance can be converted to probability

---

8. Worked Example: Stock Buy/No-Buy Classifier

Problem Setup (from textbook Figure 2.9)

Features:

- x_1 = momentum (rate of price change)
- x_2 = dividend (last quarter payment)
- x_3 = volatility (price fluctuation)

The classifier is a **plane** in 3D feature space.

The Separating Hyperplane

$$w_1 \cdot \text{momentum} + w_2 \cdot \text{dividend} + w_3 \cdot \text{volatility} + b = 0$$

Interpreting the Weights

Weight	Sign	Interpretation
w_1 (momentum)	Positive	Higher momentum favors "buy"
w_2 (dividend)	Positive	Higher dividend favors "buy"
w_3 (volatility)	Negative	Higher volatility disfavors "buy"

The **magnitude** of each weight indicates how strongly that feature influences the decision.

---

9. Higher Dimensions: The Hyperplane

Definition

A **hyperplane** in n-dimensional space is an (n-1)-dimensional subspace that divides the space into two half-spaces.

Dimension	Hyperplane	Divides space into
2D	Line (1D)	Two half-planes
3D	Plane (2D)	Two half-spaces
nD	(n-1)-dimensional hyperplane	Two half-spaces

The Beautiful Consistency

The equation $\vec{w}^T\vec{x} + b = 0$ works **identically** regardless of dimension!

---

10. Limitations of Linear Classifiers

When Hyperplanes Fail

Not all datasets can be separated by a hyperplane. Consider points arranged in concentric circles—no line can separate the inner from outer points.

Solutions

1. **Feature transformation**: Map to higher dimensions where linear separation is possible
2. **Nonlinear classifiers**: Use curved decision boundaries (covered in later chapters)
3. **Kernel methods**: Implicitly work in higher dimensions

Key Takeaways

1. **Lines and planes have elegant parametric/normal formulations** that generalize to any dimension
2. **The linear ML model $\vec{w}^T\vec{x} + b = 0$ is geometrically a hyperplane**
3. $\vec{w}$ = normal vector (orientation)
4. b = position parameter (a scalar constant derived from a point on the plane)
5. **The weight vector $\vec{w}$ is perpendicular to the decision boundary** — proven by showing $\vec{w}^T(\vec{x}_1 - \vec{x}_2) = 0$ for any two boundary points
6. **Training learns the optimal hyperplane** that best separates training data
7. **Classification uses the sign of $\vec{w}^T\vec{x} + b$** to determine which side of the hyperplane a point lies on
8. **Dot product notation**: $\vec{a} \cdot \vec{b} = \vec{a}^T\vec{b}$ — these are equivalent, but transpose notation is standard in ML
9. **This geometric view extends to all dimensions**, providing intuition even when we can't visualize the space

---

Practice Problems

Problem 1: Line Parameterization

Given points $\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$:

a) Write the parametric equation of the line

b) Find the point corresponding to α = 0.5

c) What value of α gives the point (3, 4)?

d) Is the combination with α = 1.5 a convex combination? Why or why not?

Problem 2: Hyperplane Identification

For the hyperplane $2x_1 - 3x_2 + 1 = 0$:

a) What is the normal vector (weight vector)?

b) Which side of the hyperplane contains the origin?

c) Find a point on the hyperplane.

d) Verify that the normal is perpendicular to a vector lying in the hyperplane.

Problem 3: Classifier Interpretation

A trained classifier has $\vec{w} = \begin{bmatrix} 0.5 \\ -0.3 \\ 0.8 \end{bmatrix}$ and $b = -1$.

a) Which feature has the strongest influence?

b) Classify the point $\vec{x} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$

c) Is the classifier more sensitive to increases in x_1 or x_2?

d) Calculate the signed distance from the point to the decision boundary.

Problem 4: Proving Perpendicularity

For the decision boundary $x_1 + 2x_2 - 4 = 0$:

a) Find three distinct points on this line.

b) Compute two different "within-boundary" vectors using pairs of these points.

c) Verify that $\vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is perpendicular to both vectors.

Solutions to Practice Problems

Solution 1: Line Parameterization

Given points $\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$

a) Write the parametric equation of the line

Using the formula $\vec{x} = (1-\alpha)\vec{a} + \alpha\vec{b}$:

$$\vec{x} = (1-\alpha)\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \alpha\begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

Expanding:

$$\vec{x} = \begin{bmatrix} 1 - \alpha + 5\alpha \\ 2 - 2\alpha + 6\alpha \end{bmatrix} = \begin{bmatrix} 1 + 4\alpha \\ 2 + 4\alpha \end{bmatrix}$$

Or equivalently: $\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \alpha\begin{bmatrix} 4 \\ 4 \end{bmatrix}$

b) Find the point corresponding to α = 0.5

$$\vec{x} = \begin{bmatrix} 1 + 4(0.5) \\ 2 + 4(0.5) \end{bmatrix} = \begin{bmatrix} 1 + 2 \\ 2 + 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

The point is **(3, 4)**, which is the midpoint between $\vec{a}$ and $\vec{b}$.

c) What value of α gives the point (3, 4)?

From part (b), we already found that **α = 0.5** gives (3, 4).

Alternatively, solving: $1 + 4\alpha = 3 \Rightarrow \alpha = 0.5$ ✓

d) Is the combination with α = 1.5 a convex combination? Why or why not?

When α = 1.5:

- Weight on $\vec{a}$: $(1 - 1.5) = -0.5$
- Weight on $\vec{b}$: 1.5

No, this is NOT a convex combination because one weight is negative (-0.5).

For a convex combination, we require:

1. Weights sum to 1 ✓ (−0.5 + 1.5 = 1)
2. All weights are non-negative ✗ (−0.5 < 0)

The point at α = 1.5 is $\begin{bmatrix} 1 + 4(1.5) \\ 2 + 4(1.5) \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$, which lies beyond $\vec{b}$ on the line.

Solution 2: Hyperplane Identification

For the hyperplane $2x_1 - 3x_2 + 1 = 0$

a) What is the normal vector (weight vector)?

Comparing with $\vec{w}^T \vec{x} + b = 0$, we identify:

$$\vec{w} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \quad b = 1$$

b) Which side of the hyperplane contains the origin?

Evaluate $f(\vec{x}) = \vec{w}^T\vec{x} + b$ at the origin $\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$:

$$f\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = 2(0) - 3(0) + 1 = 1 > 0$$

Since the result is **positive**, the origin is on the **positive side** of the hyperplane.

c) Find a point on the hyperplane.

We need a point where $2x_1 - 3x_2 + 1 = 0$.

Let $x_1 = 1$:

$$2(1) - 3x_2 + 1 = 0$$

$$3 - 3x_2 = 0$$

$$x_2 = 1$$

So $\vec{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is on the hyperplane.

Verification: $2(1) - 3(1) + 1 = 2 - 3 + 1 = 0$ ✓

d) Verify that the normal is perpendicular to a vector lying in the hyperplane.

First, find another point on the hyperplane. Let $x_1 = 4$:

$$2(4) - 3x_2 + 1 = 0 \Rightarrow x_2 = 3$$

So $\vec{x}_1 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ is also on the hyperplane.

Vector within the hyperplane:

$$\vec{x}_1 - \vec{x}_0 = \begin{bmatrix} 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Check perpendicularity via dot product:

$$\vec{w} \cdot (\vec{x}_1 - \vec{x}_0) = \begin{bmatrix} 2 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \end{bmatrix} = 2(3) + (-3)(2) = 6 - 6 = 0 \checkmark$$

The dot product is zero, confirming $\vec{w}$ is perpendicular to the hyperplane.

Solution 3: Classifier Interpretation

Given $\vec{w} = \begin{bmatrix} 0.5 \\ -0.3 \\ 0.8 \end{bmatrix}$ and $b = -1$.

a) Which feature has the strongest influence?

Compare the absolute values of the weights:

- $|w_1| = |0.5| = 0.5$
- $|w_2| = |-0.3| = 0.3$
- $|w_3| = |0.8| = 0.8$

Feature x_3 has the strongest influence because $|w_3| = 0.8$ is the largest.

b) Classify the point $\vec{x} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$

Compute:

$$f(\vec{x}) = \vec{w}^T \vec{x} + b = 0.5(2) + (-0.3)(1) + 0.8(3) + (-1)$$

$$= 1.0 - 0.3 + 2.4 - 1.0 = 2.1$$

Since $f(\vec{x}) = 2.1 > 0$, the point is classified as **POSITIVE CLASS**.

c) Is the classifier more sensitive to increases in x_1 or x_2?

Sensitivity is determined by the magnitude of the weights:

- $|w_1| = 0.5$ (sensitivity to x_1)
- $|w_2| = 0.3$ (sensitivity to x_2)

The classifier is **more sensitive to x_1** because $|w_1| > |w_2|$.

Additionally, note that:

- Increasing x_1 increases $f(\vec{x})$ (pushes toward positive class) since $w_1 > 0$
- Increasing x_2 decreases $f(\vec{x})$ (pushes toward negative class) since $w_2 < 0$

d) Calculate the signed distance from the point to the decision boundary.

The signed distance formula is:

$$d = \frac{\vec{w}^T\vec{x} + b}{\|\vec{w}\|}$$

First, compute $\|\vec{w}\|$:

$$\|\vec{w}\| = \sqrt{0.5^2 + (-0.3)^2 + 0.8^2} = \sqrt{0.25 + 0.09 + 0.64} = \sqrt{0.98} \approx 0.99$$

We already found $\vec{w}^T\vec{x} + b = 2.1$

Therefore:

$$d = \frac{2.1}{0.99} \approx 2.12$$

The point is approximately **2.12 units away** from the decision boundary, on the positive side.

Solution 4: Proving Perpendicularity

For the decision boundary $x_1 + 2x_2 - 4 = 0$

a) Find three distinct points on this line.

Solve for different values:

- Let $x_1 = 0$: $0 + 2x_2 = 4 \Rightarrow x_2 = 2$. Point: $\vec{p}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$
- Let $x_1 = 2$: $2 + 2x_2 = 4 \Rightarrow x_2 = 1$. Point: $\vec{p}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$
- Let $x_1 = 4$: $4 + 2x_2 = 4 \Rightarrow x_2 = 0$. Point: $\vec{p}_3 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$

Verification:

- $\vec{p}_1$: $0 + 2(2) - 4 = 0$ ✓
- $\vec{p}_2$: $2 + 2(1) - 4 = 0$ ✓
- $\vec{p}_3$: $4 + 2(0) - 4 = 0$ ✓

b) Compute two different "within-boundary" vectors using pairs of these points.

Vector 1 (from $\vec{p}_1$ to $\vec{p}_2$):

$$\vec{v}_1 = \vec{p}_2 - \vec{p}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Vector 2 (from $\vec{p}_1$ to $\vec{p}_3$):

$$\vec{v}_2 = \vec{p}_3 - \vec{p}_1 = \begin{bmatrix} 4 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$$

Note: $\vec{v}_2 = 2\vec{v}_1$, which makes sense since all three points are collinear (on the same line).

c) Verify that $\vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is perpendicular to both vectors.

Check $\vec{w} \perp \vec{v}_1$:

$$\vec{w} \cdot \vec{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 1(2) + 2(-1) = 2 - 2 = 0 \checkmark$$

Check $\vec{w} \perp \vec{v}_2$:

$$\vec{w} \cdot \vec{v}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ -2 \end{bmatrix} = 1(4) + 2(-2) = 4 - 4 = 0 \checkmark$$

Conclusion: The weight vector $\vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is perpendicular to every vector that lies within the decision boundary $x_1 + 2x_2 - 4 = 0$. This confirms that **the weight vector is the normal to the decision boundary**.

References

- Chaudhury, K. *Math and Architectures of Deep Learning*, Section 2.8
- For dot product and cosine relationship: Appendix A.1
- Equation 2.9 (dot product as transpose): Page 33